International Conference on Computational Modeling and Security (CMS2016)

Impact of Processing and Analyzing Healthcare Big Data on Cloud Computing Environment by Implementing Hadoop Cluster

Sreekanth Rallapalli[a,*], Gondkar RR[b], Uma Pavan Kumar Ketavarapu[c]

[a]Research Scholar,R&D Center,Bharathiyar University,Coiambatore,India
[b]Professor, AIT, Bangalore, India[c]Associate Professor, AIMS, Bangalore, India

Abstract

The critical challenge that the healthcare organizations are facing is to analyze the large-scale data. With the rapid growth of various healthcare applications, various devices used in healthcare generate varieties of data. The data need to be processed and effectively analyzed for better decision making. Cloud computing is a promising technology which can provide on-demand services for storage, processing and analyzing the data. The traditional data processing systems no longer has an ability to process such huge data. In order to achieve a better performance and to solve the scalability issues we need a better distributed system on cloud environment. Hadoop is a framework which can process large scale data sets on distributed environment. Hadoop can be deployed on cloud environment to process the large scale healthcare data. Healthcare applications are being supplied through internet and cloud services rather than using as traditional software. Healthcare providers need to have real time information to provide quality healthcare.This paper discuss on the impacts of data processing and analyzing large scale healthcare data on cloud computing environment.

Keywords: Big Data;Cloud Computing;Cluster;Hadoop;Healthcare.

1. Introduction

Healthcare providers across the world are migrating to use the software applications as a service which is provided by most of the cloud providers. As Healthcare information is confidential and to provide better quality healthcare, various stakeholders should exchange the patient information, clinical information in a secured way. As healthcare data is being available in large data sets and in various formats, cloud environment is the efficient way to store and process the data. Today most of the software applications are being deployed in the data centers[1]. In order to perform complex computations cloud computing is a dominant architecture which can efficiently perform large-scale data computation by providing the scalable resources[2]. Leading research organizations like Gartner Inc and market intelligence firms like IDC reports that Big data and Cloud computing are emerging technologies in today's Business Intelligence market.

*Corresponding author. SreekanthRallapalli, Tel:00-91-9008716563
E-mail:rsreekanth1@yahoo.com

As per IBM reports everyday 2.5 quintillion bytes of data is created and 90% of data is created in the past 2 years[3]. As per McKinsey analysis Big data is changing the paradigm of health care and new insights of the data is creating new value for the payers and providers[4]. So far health care data is available in form of Electronic Medical Records (EMR), Electronic Health Records (EHR), and Patient Medical Records (PMR). The collection of digitized medical records is increased over a period of time created large data sets. The large data sets of health care need to be effectively analyzed and try to resolve the various problems the industry is facing. Quality health care, reduction in medical cost, efficient decision making to provide the appropriate healthcare, finding patterns for unnecessary hospital readmissions are some of the issues can be resolved by Big data technology. New value from the large data sets can be analyzed by building effective analytical tools that assist patients, physicians and various stake holders in health care. Cloud computing has resolved most of the health care data related issues such as standardization of exchange of health care records, privacy, and network security. Security is the major issue while sharing the Health care data in the clouds. Access control mechanism which will support sharing of the EHR[5] will address the issue.

Hadoop framework solves most of the problems related to huge data processing. Hadoop applications run on clusters which are built using commodity hardware. One of the important features of Hadoop is fault tolerant. MapReduce is the computing paradigm used in Hadoop as it provides Hadoop Distributed File System (HDFS) which stores data on the nodes. Healthcare data sets need to be analyzed on the hadoop clusters in cloud computing environment to solve various issues of the industry.

2. Sharing of Healthcare Data on Clouds

Healthcare providers across the globe are now willing to move the data to the cloud for reducing the operational costs. Adoption of cloud computing is still not popular in many countries due to various challenges like security, authentication and access control mechanisms. Researchers have provided various solutions to overcome the challenges[5, 6].Many of the physicians across the globe do not have proper information while dealing with patient. The patient needs to carry all his past history records and then explain about his past medical data. This sometimes led to inaccurate decision based on patient records. Various organizations have developed integrated solutions to meet the above said challenges. But it requires a new infrastructure to develop in the healthcare organizations[7].

To overcome the issues of security it is proposed to have a segregated network for Electronic Health care Records[8]. Also it is recommended to check for vulnerabilities, providing data loss prevention programs, installation of firewalls, and to provide all national level policies on healthcare.

3. Big Data for Healthcare

Big data in today's world is identified by 5 V's, Volume, Velocity, Variety, Veracity and Value[9]. As discussed earlier healthcare is one of the organizations which are generating huge data sets. Analysis of large sets of patient records involve in identifying the patient clusters and correlations. This also involves in developing various models using statistical principles or machine learning techniques[10, 11]. Confidentiality of the healthcare data is one of the major problem the organizations are facing where the healthcare provider need to protect. As healthcare generates complex data base the applications developed in big data should handle such data complex in nature. Big data applications provide solutions for computational bio medicine, genomic data,